

<sup>1</sup> Fitness tracking reveals task-specific associations  
<sup>2</sup> between memory, mental health, and exercise

<sup>3</sup> Jeremy R. Manning<sup>1, \*</sup>, Gina M. Notaro<sup>1</sup>, Esme Chen<sup>1</sup>, and Paxton C. Fitzpatrick<sup>1</sup>

<sup>4</sup> <sup>1</sup>Dartmouth College, Hanover, NH

<sup>5</sup> \*Address correspondence to jeremy.r.manning@dartmouth.edu

<sup>6</sup>

March 8, 2022

7                   **Abstract**

8         Physical exercise can benefit both physical and mental well-being. Different forms of exercise  
9         (e.g., aerobic versus anaerobic; running versus walking, swimming, or yoga; high-intensity  
10       interval training versus endurance workouts; etc.) impact physical fitness in different ways.  
11       For example, running may substantially impact leg and heart strength but only moderately  
12       impact arm strength. We hypothesized that the mental benefits of exercise might be similarly  
13       differentiated. We focused specifically on how different forms of exercise might relate to different  
14       aspects of memory and mental health. To test our hypothesis, we collected (in aggregate) roughly  
15       a century's worth of fitness data. We then asked participants to fill out surveys asking them to  
16       self-report on different aspects of their mental health. We also asked participants to engage in  
17       a battery of memory tasks that tested their short and long term episodic, semantic, and spatial  
18       memory performance. We found that participants with similar exercise habits and fitness profiles  
19       tended to also exhibit similar mental health and task performance profiles. These effects were  
20       task-specific in that different exercise patterns or fitness characteristics varied with different  
21       aspects of memory, on different tasks. Taken together, these findings provide foundational work  
22       for designing exercise interventions that target specific components of cognitive performance  
23       and mental health by leveraging low-cost fitness tracking devices.

24                   **Introduction**

25       Engaging in physical activity (exercise) can improve physical fitness by increasing muscle strength  
26       [9, 18, 21, 30], bone density [1, 8, 19], cardiovascular performance [22, 28], lung capacity [20,  
27       although see [31]], and endurance [39]. Exercise can also improve mental health [2, 4, 10, 23, 27,  
28       29, 36] and cognitive performance [2, 3, 6, 11].

29       The physical benefits of exercise can be explained by stress-responses of the affected body  
30       tissues. For example, skeletal muscles that are taxed during exercise exhibit stress responses [24]  
31       that can in turn affect their growth or atrophy [32]. By comparison, the benefits of exercise on mental  
32       health are less direct. For example, one hypothesis is that exercise leads to specific physiological  
33       changes, such as increased aminergic synaptic transmission and endorphin release, which in turn

34 act on neurotransmitters in the brain [27]. Speculatively, if different exercise regimens lead to  
35 different neurophysiological responses, one might be able to map out a spectrum of signalling and  
36 transduction pathways that are impacted by a given type, duration, and intensity of exercise in  
37 each brain region. For example, prior work has shown that exercise increases acetylcholine levels,  
38 starting in the vicinity of the exercised muscles [33]. Acetylcholine is thought to play an important  
39 role in memory formation [e.g., by modulating specific synaptic inputs from entorhinal cortex to  
40 the hippocampus, albeit in rodents; 26]. Given the central role that these medial temporal lobe  
41 structures play in memory, changes in acetylcholine might lead to specific changes in memory  
42 formation and retrieval.

43 In the present study, we hypothesize that (a) different exercise regimens will have different,  
44 quantifiable impacts on cognitive performance and mental health, and that (b) these impacts will  
45 be consistent across individuals. To this end, we collected a year of real-world fitness tracking  
46 data from each of 113 participants. We then asked each participant to fill out a brief survey in  
47 which they self-evaluated and self-reported several aspects of their mental health. Finally, we ran  
48 each participant through a battery of memory tasks, which we used to evaluate their memory  
49 performance along several dimensions. We searched the data for potential associations between  
50 memory, mental health, and exercise.

## 51 **Methods**

52 We ran an online experiment using the Amazon Mechanical Turk (MTurk) platform [12]. We col-  
53 lected data about each participant's fitness and exercise habits, a variety of self-reported measures  
54 concerning their mental health, and about their performance on a battery of memory tasks.

55 **Experiment**

56 **Participants**

57 We recruited experimental participants by posting our experiment as a Human Intelligence Task  
58 (HIT) on the MTurk platform. We limited participation to MTurk Workers who had been assigned  
59 a “master worker” designation on the platform, given to workers who score highly across several  
60 metrics on a large number of HITs, according to a proprietary algorithm managed by Amazon.  
61 One criteria embedded into the algorithm is a requirement that master workers must maintain a  
62 HIT acceptance rate of at least 95%. We further limited our participant pool to participants who  
63 self-reported that they were fluent in English and regularly used a Fitbit fitness tracker device.  
64 A total of 160 workers accepted our HIT in order to participate in our experiment. Of these,  
65 we excluded all participants who failed to log into their Fitbit account (giving us access to their  
66 anonymized fitness tracking data), encountered technical issues (e.g., by accessing the HIT using an  
67 incompatible browser, device, or operating system), or who ended their participation prematurely,  
68 before completing the full study. In all, 113 participants contributed usable data to the study.

69 For their participation, workers received a base payment of \$5 per hour (computed in 15  
70 minute increments, rounded up to the nearest 15 minutes), plus an additional performance-based  
71 bonus of up to \$5. Our recruitment procedure and study protocol were approved by Dartmouth’s  
72 Committee for the Protection of Human Subjects. We obtained informed consent using an online  
73 form administered to all prospective participants prior to enrolling them in our study. All methods  
74 were performed in accordance with the relevant guidelines and regulations.

75 **Gender, age, and race.** Of the 113 participants who contributed usable data, 77 reported their  
76 gender as female, 35 as male, and 1 chose not to report their gender. Participants ranged in age  
77 from 19–68 years old (25<sup>th</sup> percentile: 28.25 years; 50<sup>th</sup> percentile: 32 years; 75<sup>th</sup> percentile: 38  
78 years). Participants reported their race as White (90 participants), Black or African American (11  
79 participants), Asian (7 participants), Other (4 participants), and American Indian or Alaska Native  
80 (3 participants). One participant opted not to report their race.

81 **Languages.** All participants reported that they were fluent in either 1 and 2 languages (25<sup>th</sup>  
82 percentile: 1; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 1), and that they were “familiar” with between 1  
83 and 11 languages (25<sup>th</sup> percentile: 1; 50<sup>th</sup> percentile: 2; 75<sup>th</sup> percentile: 3).

84 **Reported medical conditions and medications.** Participants reported having and/or taking med-  
85 ications pertaining to the following medical conditions: anxiety or depression (4 participants),  
86 recent head injury (2 participants), high blood pressure (1 participant), bipolar disorder (1 partici-  
87 pant), hypothyroidism (1 participant), and other unspecified conditions or medications (1 partici-  
88 pant). Participants reported their current and typical stress levels on a Likert scale as very relaxed  
89 (-2), a little relaxed (-1), neutral (0), a little stressed (1), or very stressed (2). The “current” stress  
90 level reflected participants’ stress at the time they participated in the experiment. Their responses  
91 ranged from -2 to 2 (current stress: 25<sup>th</sup> percentile: -2; 50<sup>th</sup> percentile: -1; 75<sup>th</sup> percentile: 1; typical  
92 stress: 25<sup>th</sup> percentile: 0; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 1). Participants also reported their  
93 current level of alertness on a Likert scale as very sluggish (-2), a little sluggish (-1), neutral (0), a  
94 little alert (1), or very alert (2). Their responses ranged from -2 to 2 (25<sup>th</sup> percentile: 0; 50<sup>th</sup> per-  
95 centile: 1; 75<sup>th</sup> percentile: 2). Nearly all (111 out of 113) participants reported that they had normal  
96 color vision, and 15 participants reported uncorrected visual impairments (including dyslexia and  
97 uncorrected near- or far-sightedness).

98 **Residence and level of education.** Participants reported their residence as being located in the  
99 suburbs (36 participants), a large city (30 participants), a small city (23 participants), rural (14 partic-  
100 ipants), or a small town (10 participants). Participants reported their level of education as follows:  
101 College graduate (42 participants), Master’s degree (23 participants), Some college (21 partici-  
102 pants), High school graduate (9 participants), Associate’s degree (8 participants), Other graduate  
103 or professional school (5 participants), Some graduate training (3 participants), or Doctorate (2  
104 participants).

105 **Reported water and coffee intake.** Participants reported the number of 8 oz cups of water and  
106 coffee they had consumed prior to accepting the HIT. Water consumption ranged from 0–6 cups

<sup>107</sup> (25<sup>th</sup> percentile: 1; 50<sup>th</sup> percentile: 3; 75<sup>th</sup> percentile: 4). Coffee consumption ranged from 0–4 cups  
<sup>108</sup> (25<sup>th</sup> percentile: 0; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 2).

<sup>109</sup> **Tasks**

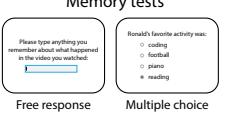
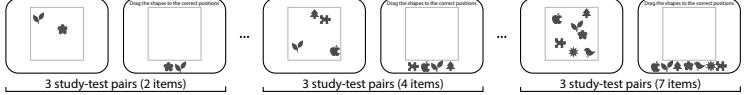
<sup>110</sup> Upon accepting the HIT posted on MTurk, each worker was directed to read and fill out a screening  
<sup>111</sup> and consent form, and to share access to their anonymized Fitbit data via their Fitbit account. After  
<sup>112</sup> consenting to participate in our study and successfully sharing their Fitbit data, participants filled  
<sup>113</sup> out a survey and then engaged in a series of memory tasks (Fig. 1). All stimuli and code for running  
<sup>114</sup> the full MTurk experiment may be found [here](#).

<sup>115</sup> **Survey questions.** We collected the following demographic information from each participant:  
<sup>116</sup> their birth year, gender, highest (academic) degree achieved, race, language fluency, and language  
<sup>117</sup> familiarity. We also collected information about participants' health and wellness, including about  
<sup>118</sup> their vision, alertness, stress, sleep, coffee and water consumption, location of their residence,  
<sup>119</sup> activity typically required for their job, and exercise habits.

<sup>120</sup> **Free recall (Fig. 1a).** Participants studied a sequence of four word lists, each comprising 16 words.  
<sup>121</sup> After studying each list, participants received an immediate memory test, whereby they were asked  
<sup>122</sup> to type (one word at a time) any words they remembered from the just-studied list, in any order.

<sup>123</sup> Words were presented for 2 s each, in black text on a white background, followed by a 2 s blank  
<sup>124</sup> (white) screen. After the final 2 s pause, participants were given 90 s to type in as many words  
<sup>125</sup> as they could remember, in any order. The memory test was constructed such that the participant  
<sup>126</sup> could only see the text of the current word they were typing; when they pressed any non-letter  
<sup>127</sup> key, the current word was submitted and the text box they were typing in was cleared. This was  
<sup>128</sup> intended to prevent participants from retroactively editing their previous responses.

<sup>129</sup> The word lists participants studied were drawn from the categorized lists reported by [40]. Each  
<sup>130</sup> participant was assigned four unique randomly chosen lists (in a randomized order), selected from  
<sup>131</sup> a full set of 16 lists. Each chosen list was then randomly shuffled before presenting the words to

	Main task and immediate memory test				Delayed memory test
a.	1 Free recall	Study words 	Memory test 		5 
b.	2 Naturalistic recall	Watch a short video (The Temple of Knowledge) 	Memory tests 		6 
c.	3 Foreign language flashcards	Study flashcards 	Memory test 		7 
d.	4 Spatial learning	Memorize the positions of increasing numbers of shapes 			N/A

**Figure 1: Battery of memory tasks.** **a. Free recall.** Participants study 16 words (presented one at a time), followed by an immediate memory test where they type each word they remember from the just-studied list. In the delayed memory test, participants type any words they remember studying, from any list. **b. Naturalistic recall.** Participants watch a brief video, followed by two immediate memory tests. The first test asks participants to write out what happened in the video. The second test has participants answer a series of multiple choice questions about the conceptual content of the video. In the delayed memory test, participants (again) write out what happened in the video. **c. Foreign language flashcards.** Participants study a sequence of 10 English-Gaelic word pairs, each presented with an illustration of the given word. During an immediate memory test, participants perform a multiple choice test where they select the Gaelic word that corresponds to the given photograph. During the delayed memory test, participants perform a second multiple choice test, where they select the Gaelic word that corresponds to each of a new set of photographs. **d. Spatial learning.** In each trial, participants study a set of randomly positioned shapes. Next, the shapes' positions are altered, and participants are asked to drag the shapes back to their previous positions. **All panels.** The gray numbers denote the order in which participants experienced each task or test.

<sup>132</sup> the participants. Participants also performed a final delayed memory test where they were given  
<sup>133</sup> 180 s to type out any words they remembered from *any* of the 4 lists they had studied.

<sup>134</sup> Recalled words within an edit distance of 2 (i.e., a Levenshtein Distance less than or equal to  
<sup>135</sup> 2) of any word in the wordpool were “autocorrected” to their nearest match. We also manually  
<sup>136</sup> corrected clear typos or misspellings by hand (e.g., we corrected “hippoptumas” to “hippopota-  
<sup>137</sup> mus”, “zucinni” to “zucchini”, and so on). Finally, we lemmatized each submitted word to match  
<sup>138</sup> the plurality of the matching wordpool word (e.g., “bongo” was corrected to “bongos”, and so  
<sup>139</sup> on). After applying these corrections, any submitted words that matched words presented on the  
<sup>140</sup> just-studied list were tagged as “correct” recalls, and any non-matching words were discarded  
<sup>141</sup> as “errors.” Because participants were not allowed to edit the text they entered, we chose not to  
<sup>142</sup> analyze these putative “errors,” since we could not distinguish typos from true misrememberings.

<sup>143</sup> **Naturalistic recall (Fig. 1b).** Participants watched a 2.5 minute video clip entitled “The Temple  
<sup>144</sup> of Knowledge.” The video comprises an animated story told to StoryCorps by Ronald Clark, who  
<sup>145</sup> was interviewed by his daughter, Jamilah Clark. The narrator (Ronald) discusses growing up  
<sup>146</sup> living in an apartment over the Washington Heights branch of the New York Public Library, where  
<sup>147</sup> his father worked as a custodian during the 1940s.

<sup>148</sup> After watching the video clip, participants were asked to type out anything they remembered  
<sup>149</sup> about what happened in the video. They typed their responses into a text box, one sentence at a  
<sup>150</sup> time. When the participant pressed the return key or typed any final punctuation mark (“.”, “!”, or  
<sup>151</sup> “?”) the text currently entered into the box was “submitted” and added to their transcript, and the  
<sup>152</sup> text box was cleared to prevent further editing of any already-submitted text. This was intended to  
<sup>153</sup> prevent participants from retroactively editing their previous responses. Participants were given  
<sup>154</sup> up to 10 minutes to enter their responses. After 4 minutes, participants were given the option of  
<sup>155</sup> ending the response period early, e.g., if they felt they had finished entering all of the information  
<sup>156</sup> they remembered. Each participant’s transcript was constructed from their submitted responses by  
<sup>157</sup> combining the sentences into a single document and removing extraneous whitespace characters.  
<sup>158</sup> Following this 4–10 minute free response period, participants were given a series of 10 multiple

<sup>159</sup> choice questions about the conceptual content of the story. All participants received the same  
<sup>160</sup> questions, in the same order. Participants also performed a final delayed memory test, where they  
<sup>161</sup> carried out the free response recall task a second time, near the end of the testing session. This  
<sup>162</sup> resulted in a second transcript, for each participant.

<sup>163</sup> **Foreign language flashcards (Fig. 1c).** Participants studied a series of 10 English-Gaelic word  
<sup>164</sup> pairs in a randomized order. We selected the Gaelic language both for its relatively small number  
<sup>165</sup> of native speakers and for its dissimilarity to other commonly spoken languages amongst MTurk  
<sup>166</sup> workers. We verified (via self report) that all of our participants were fluent in English and that  
<sup>167</sup> they were neither fluent nor familiar with Gaelic.

<sup>168</sup> Each word's "flashcard" comprised a cartoon depicting the given word, the English word or  
<sup>169</sup> phrase in lowercase text (e.g., "the boy"), and the Gaelic word or phrase in uppercase text (e.g.,  
<sup>170</sup> "BUACHAILL"). Each flashcard was displayed for 4 s, followed by a 3 s interval (during which  
<sup>171</sup> the screen was cleared) prior to the next flashcard presentation.

<sup>172</sup> After studying all 10 flashcards, participants were given a multiple choice memory test where  
<sup>173</sup> they were shown a series of novel photographs, each depicting one of the 10 words they had  
<sup>174</sup> learned. They were asked to select which (of 4 unique options) Gaelic word went with the given  
<sup>175</sup> picture. The 3 incorrect options were selected at random (with replacement across trials), and the  
<sup>176</sup> order in which the choices appeared to the participant were also randomized. Each of the 10 words  
<sup>177</sup> they had learned were tested exactly once.

<sup>178</sup> Participants also performed a final delayed memory test, where they were given a second set of  
<sup>179</sup> 10 questions (again, one per word they had studied). For this second set of questions participants  
<sup>180</sup> were prompted with a new set of novel photographs, and new randomly chosen incorrect choices  
<sup>181</sup> for each question. Each of the 10 original words they had learned were (again) tested exactly once  
<sup>182</sup> during this final memory test.

<sup>183</sup> **Spatial learning (Fig. 1d).** Participants performed a series of study-test trials where they memo-  
<sup>184</sup> rized the onscreen spatial locations of two or more shapes. During the study phase of each trial,

185 a set of shapes appeared on the screen for 10 s, followed by 2 s of blank (white) screen. During the  
186 test phase of each trial, the same shapes appeared onscreen again, but this time they were vertically  
187 aligned and sorted horizontally in a random order. Participants were instructed to drag (using the  
188 mouse) each shape to its studied position, and then to click a button to indicate that the placements  
189 were complete.

190 In different study-test trials, participants learned the locations of different numbers of shapes  
191 (always drawn from the same pool of 7 unique shapes, where each shape appeared at most one  
192 time per trial). They first performed three trials where they learned the locations of 2 shapes; next  
193 three trials where they learned the locations of 3 shapes; and so on until their last three trials, where  
194 (during each trial) they learned the locations of 7 shapes. All told, each participant performed 18  
195 study-test trials of this spatial learning task (3 trials for each of 2, 3, 4, 5, 6, and 7 shapes).

#### 196 **Fitness tracking using Fitbit devices**

197 To gain access to our study, participants provided us with access to all data associated with their  
198 Fitbit account from the year (365 calendar days) up to and including the day they accepted the HIT.  
199 We filtered out all identifiable information (e.g., participant names, GPS coordinates, etc.) prior to  
200 importing their data.

#### 201 **Collecting and processing Fitbit data**

202 The fitness tracking data associated with participants' Fitbit accounts varied in scope and duration  
203 according to which device the participant owned (Fig. S1), how often the participant wore (and/or  
204 synced) their tracking device, and how long they had owned their device. For example, while all  
205 participants' devices supported basic activity metrics such as daily step counts, only a subset of  
206 the devices with heart rate monitoring capabilities provided information about workout intensity,  
207 resting heart rate, and other related measures. Across all devices, we collected the following infor-  
208 mation: heart rate data, sleep tracking data, logged bodyweight measurements, logged nutrition  
209 measurements, Fitbit account and device settings, and activity metrics.

210 **Heart rate.** If available, we extracted all heart rate data collected by participants' Fitbit device(s)  
211 and associated with their Fitbit profile. Depending on the specific device model(s) and settings, this  
212 included second-by-second, minute-by-minute, daily summary, weekly summary, and/or monthly  
213 summary heart rate information. These summaries include information about participants' aver-  
214 age heart rates, and the amount of time they were estimated to have spent in different "heart rate  
215 zones" (rest, out-of-range, fat burn, cardio, or peak, as defined by their Fitbit profile), as well as an  
216 estimate of the number of estimated calories burned while in each heart rate zone.

217 **Sleep.** If available, we extracted all sleep data collected by participants' Fitbit device(s). Depend-  
218 ing on the specific device model(s) and settings, this included nightly estimates of the duration  
219 and quality of sleep, as well as the amount of time spent in each sleep stage (awake, REM, light, or  
220 deep).

221 **Weight.** If available, we extracted any weight-related information affiliated with participants'  
222 Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific device  
223 model(s) and settings, this included their weight, body mass index, and/or body fat percentage.

224 **Nutrition.** If available, we extracted any nutrition-related information affiliated with participants'  
225 Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific account  
226 settings and usage behaviors, this included a log of the specific foods they had eaten (and logged)  
227 over the past year, and the amount of water consumed (and logged) each day.

228 **Account and device settings.** We extracted any settings associated with participants' Fitbit ac-  
229 counts to determine (a) which device(s) and model(s) are associated with their Fitbit account, (b)  
230 time(s) when their device(s) were last synced, and (c) battery level(s).

231 **Activity metrics.** If available, we extracted any activity-related information affiliated with par-  
232 ticipants' Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific  
233 device model(s) and settings, this included: daily step counts; daily amount of time spent in each

234 activity level (sedentary, lightly active, fairly active, or very active, as defined by their account  
235 settings and preferences); daily number of floors climbed; daily elevation change; and daily total  
236 distance traveled.

237 **Comparing recent versus baseline measurements.**

238 We were interested in separating out potential associations between *absolute* fitness metrics and  
239 *relative* metrics. To this end, in addition to assessing potential raw (absolute) fitness metrics, we  
240 also defined a simple measure of recent changes in those metrics, relative to a baseline:

$$\Delta_{R,B}m = \frac{B \sum_{i=1}^R m(i)}{R \sum_{i=R+1}^{R+B} m(i)},$$

241 where  $m(i)$  is the value of metric  $m$  from  $i - 1$  days prior to testing (e.g.,  $m(1)$  represents the value  
242 of  $m$  on the day the participant accepted the HIT, and  $m(10)$  represents the value of  $m$  9 days prior  
243 to accepting the HIT. We set  $R = 7$  and  $B = 30$ . In other words, to estimate recent changes in any  
244 metric  $m$ , we divided the average value of  $m$  taken over the prior week by the average value of  $m$   
245 taken over the 30 days before that.

246 **Exploratory correlation analyses**

247 We used a bootstrap procedure to identify reliable correlations between different memory-related,  
248 fitness-related, and demographic-related variables. For each of  $N = 10,000$  iterations, we selected  
249 (with replacement) a sample of 113 participants to include. This yielded, for each iteration, a  
250 sampled “data matrix” with one row per sampled participant and one column for each measured  
251 variable. When participants were sampled multiple times in a given iteration, as was often the  
252 case, this matrix contained duplicate rows. Next, we computed the Pearson’s correlation between  
253 each pair of columns. This yielded, for each pair of columns, a distribution of  $N$  bootstrapped  
254 correlation coefficients. If 97.5% or fewer of the coefficients for a given pair of columns had the  
255 same sign, we excluded the pair from further analysis and considered the expected correlation  
256 between those columns to be undefined. If > 97.5% of the coefficients for a given pair of columns

had the same sign (corresponding to a bootstrap-estimated two-tailed  $p$  threshold of 0.05), we computed the expected correlation coefficient as:

$$\mathbb{E}_{i,j}[r] = \tanh\left(\frac{1}{N} \sum_{n=1}^N \tanh^{-1}(\text{corr}(m(i)_n, m(j)_n))\right),$$

where  $m(x)_n$  represents column  $x$  of the bootstrapped data matrix for iteration  $n$ ,  $\tanh$  is the hyperbolic tangent, and  $\tanh^{-1}$  is the inverse hyperbolic tangent. We estimated the corresponding  $p$ -values for these correlations as one minus the proportion of bootstrapped correlations with the same sign, multiplied by two.

### Reverse correlation analyses

We sought to characterize potential associations between the *dynamics* of participants' fitness-related activities leading up to the time they participated in a memory task and their performance on the given task. For each fitness-related variable, we constructed a timeseries matrix whose rows corresponded to timepoints (sampled once per day) leading up to the day the participant accepted the HIT for our study, and whose columns corresponded to different participants. These matrices often contained missing entries, since different participants' Fitbit devices tracked fitness-related activities differently. For example, participants whose Fitbit devices lacked heart rate sensors would have missing entries for any heart rate-related variables. Or, if a given participant neglected to wear their fitness tracker on a particular day, the column corresponding to that participant would have missing entries for that day. To create stable estimates, we smoothed the timeseries of each fitness measure using a sliding window of 1 week. In other words, for each fitness measure, we replaced the "observed value" for each day with the average values of that measure (when available) over the 7 day interval ending on the given day.

In addition to this set of matrices storing timeseries data for each fitness-related variable, we also constructed a memory performance matrix,  $M$ , whose rows corresponded to different memory-related variables, and whose columns corresponded to different participants. For example, one row of the memory performance matrix reflected the average proportion of words (across lists)

281 that each participant remembered during the immediate free recall test, and so on.

282 Given a fitness timeseries matrix,  $F$ , we computed the weighted average and weighted standard  
283 error of the mean of each row of  $F$ , where the weights were given by a particular memory-related  
284 variable (row of  $M$ ). For example, if  $F$  contained participants' daily step counts, we could use  
285 any row of  $M$  to compute a weighted average across any participants who contributed step count  
286 data on each day. Choosing a row of  $M$  that corresponded to participants' performance on the  
287 naturalistic recall task would mean that participants who performed better on the naturalistic recall  
288 task would contribute more to the weighted average timeseries of daily step counts. Specifically,  
289 for each row,  $t$ , of  $F$ , we computed the weighted average (across the  $S$  participants) as:

$$\bar{f}(t) = \sum_{s=1}^S \hat{m}(s)F(t,s),$$

290 where  $\hat{m}$  denotes the normalized min-max scaling of  $m$  (the row of  $M$  corresponding to the chosen  
291 memory-related variable):

$$\hat{m} = \frac{m}{\sum_{s=1}^S \hat{m}(s)},$$

292 where

$$\hat{m} = (1 - \epsilon) \frac{m - \min(m)}{\max(m) - \min(m)} + \epsilon.$$

293 Here,  $\epsilon$  provides a lower bound on the influence of the lowest-weighted participant's data. We  
294 defined  $\epsilon = 0.001$ , ensuring that the lowest-weighted participant had relatively low (but non-zero)  
295 influence. We computed the weighted standard error of the mean as:

$$\text{SEM}_m(f(t)) = \frac{\left| \sum_{s=1}^S (F(t,s) - \bar{f}(t)) \right|}{\sqrt{S}}.$$

296 When a given row of  $F$  was missing data from one or more participants, those participants were  
297 excluded from the weighted average for the corresponding timepoint and the weights (across all  
298 remaining participants) were re-normalized to sum to 1. The above procedure yielded, for each  
299 memory variable, a timeseries of weighted average (and weighted standard error of the mean)

300 fitness tracking values leading up to the day of the experiment.

## 301 Results

302 Before testing our main hypotheses, we examined the behavioral data from each of four memory  
303 tasks: a random word list learning “free recall” task (Fig. 1); a naturalistic recall task whereby par-  
304 ticipants watched a short video and then recounted the narrative; a foreign language “flashcards”  
305 task; and a spatial learning task. Each of the first three tasks (free recall, naturalistic recall, and the  
306 flashcards task) included both an immediate (short term) memory test and a delayed (long term)  
307 memory test. The spatial learning task included only an immediate test. Participants in all four  
308 tasks exhibited general trends and tendencies that have been previously reported in prior work.  
309 We were also interested in characterizing the variability in task performance across participants.  
310 For example, if all participants exhibited near-identical behaviors or performance on a given task,  
311 we would be unable to identify how memory performance on that task varied with mental health  
312 or exercise.

313 When participants engaged in free recall of random word lists, they displayed strong primacy  
314 and recency effects [25] on the immediate memory tests (as reflected by improved memory for  
315 early and late list items; Fig. 2a, left and right panels). On the delayed memory test, the recency  
316 effect was substantially diminished (Fig. 3a, left and right panels), consistent with myriad previous  
317 studies [for review see 16]. Participants also tended to cluster their recalls according to the words'  
318 study positions [15] on both the immediate (Fig. 2a, middle panel) and delayed (Fig. 3a, middle  
319 panel) memory tests.

320 When participants engaged in naturalistic recall by recounting the narrative of a short story  
321 video, they reliably and accurately remembered the major narrative events on both the immediate  
322 (Fig. 2b) and delayed (Fig. 3b) tests. This is consistent with prior work showing that memory for  
323 rich narratives is both detailed and accurate [7, 13].

324 Performance on the foreign language flashcards task (immediate: Fig. 2c; delayed: Fig. 3c)  
325 varied substantially across participants, and did not show any clear serial position effects. Partic-

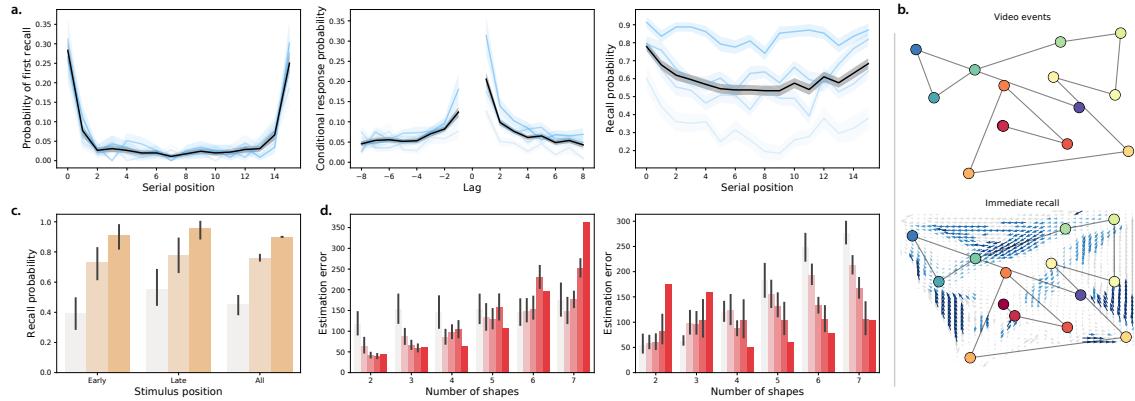
326 participants also displayed substantial variation in performance on the spatial learning task (Fig. 2d).  
327 In general, participants reported the shape's positions more accurately when there were fewer  
328 shapes. However, both the baseline estimation accuracy and the rate of decrease in accuracy as a  
329 function of increasing number of memorized locations varied substantially across participants.

330 In addition to observing substantial across-participant variability in memory performance,  
331 we also observed substantial variability in participants' fitness and activity metrics (Fig. 4). We  
332 examined recent measurements, averaged over the week prior to testing (Fig. 4a), baselined mea-  
333 surements (average over the prior week, divided by the average over the preceding 30 days;  
334 Fig. 4b), along with more gradually varying measures that tended to remain relatively static over  
335 timescales of weeks to months (Fig. 4c). Figure S6 displays across-participant distributions for  
336 a broad selection of these measures, and Figures S7, S8, S9, and S10 show different participants'  
337 fitness metrics, broken down by their performance on different memory tasks.

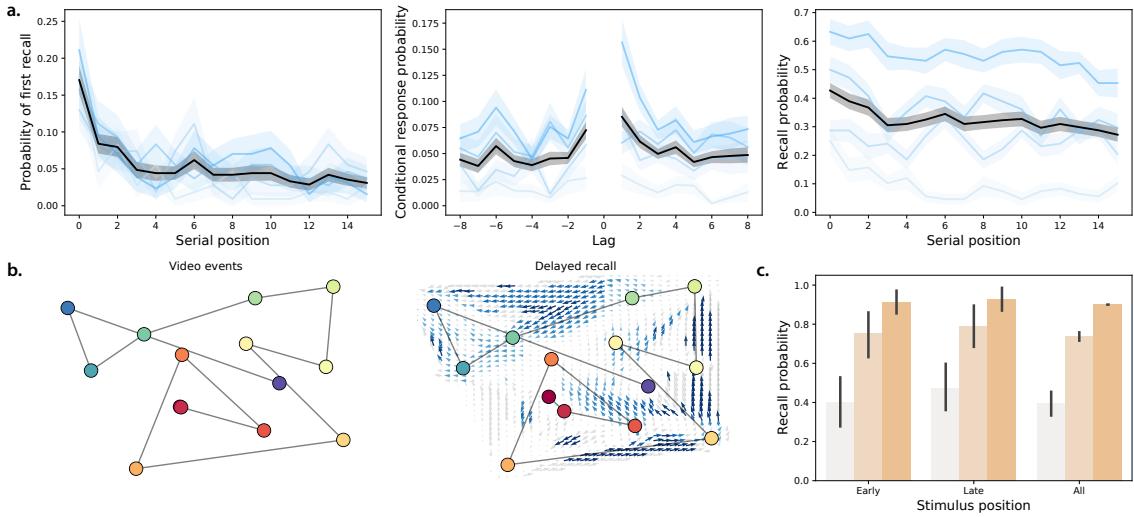
338 We wondered about potential links between the different aspects of participants' data. For  
339 example, if people who exercised in a particular way also tended to perform better on a given  
340 memory task, this could suggest that either (a) some property intrinsic to participants who exercised  
341 in a particular way might also affect their memory performance on the given task, and/or (b)  
342 particular exercise behaviors could have a causal impact on memory performance. We carried  
343 out an exploratory analysis whereby we used a bootstrap-based approach to identify reliable  
344 correlations between different aspects of memory performance (Fig. S11), different aspects of  
345 fitness (Fig. S12), different demographic attributes (Fig. S13), and correlations between memory  
346 performance, fitness information, and demographic attributes (Fig. S14). Specifically, we sought  
347 to identify correlations that were present in the same direction (i.e., positive or negative) across  
348 different subsets of participants. For each test, we report the average correlation (taken across  
349 10,000 subsets of participants, chosen with replacement) and an associated two-tailed  $p$ -value,  
350 estimated as

$$p = 2 \times (1 - q),$$

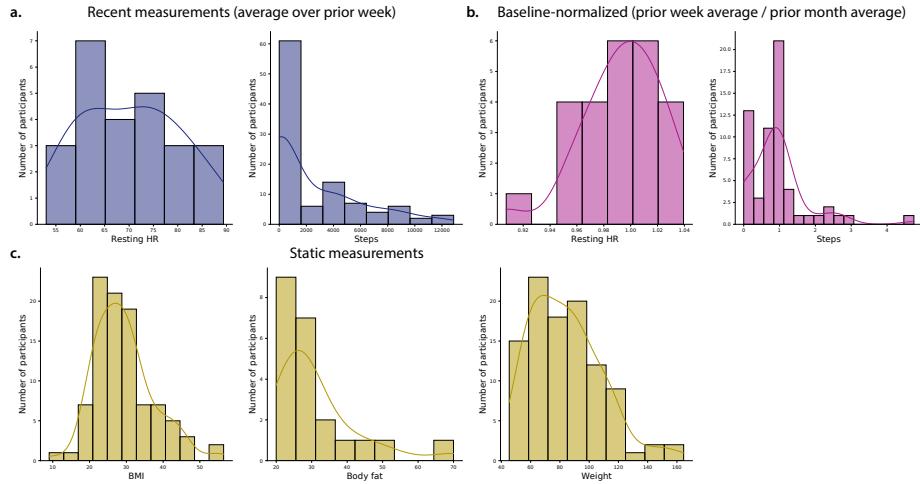
351 where  $q$  is the proportion of those 10,000 subsets that exhibited correlations in the same direction



**Figure 2: Immediate memory tests.** **a. Free recall.** Left: probability of recalling each word first as a function of its presentation position. Middle: probability of transitioning between successively recalling the word presented at position  $i$ , followed by word presented at position  $i + \text{Lag}$ . Right: probability of recalling each word as a function of its presentation position. See Figure S2 for additional details. **b. Naturalistic recall.** Top: 2D embedding of a 2.5 min video clip; each dot reflects a narrative event (red denotes early events and blue denotes later events). Bottom: 2D embedding of the averaged transcripts of participants' recounts of the narrative (dots: same format as top panel). The arrows denote the average trajectory directions through the corresponding region of text embedding space, for any participants whose recounts passed through that region. Blue arrows denote statistically reliable agreement across participants ( $p < 0.05$ , corrected). See Figure S3 for additional details. **c. Foreign language flashcards.** Each bar denotes the average proportion of correctly recalled Gaelic-English word pairs from early (first 3), late (last 3), or all (i.e., all 10) study positions. See Figure S4 for additional details. **d. Spatial learning.** Average estimation error in shape locations as a function of the number of shapes. See Figure S5 for additional details. All panels: error bars and error ribbons denote bootstrap-estimated 95% confidence intervals. Shading (saturation) denotes results for different subsets of participants assigned based on their task performance (Figs. S2, S3, S4, and S5 provide information about which performance metrics and values the shading reflects; in general more saturated colors denote participants who performed better on the given task.) In Panel d, participants are grouped in two ways; in the left panel, participants are grouped according to the  $y$ -intercepts of regression lines (estimation error as a function of the number of shapes); in the right panel, participants are grouped according to the slopes of the same regression lines.



**Figure 3: Delayed memory tests.** **a. Free recall.** These panels are in the same format as Figure 2a, but they reflect performance on the delayed free recall task. For additional details see Figure S2. **b. Naturalistic recall.** These panels are in the same format as Figure 2b, but the right panel reflects performance on the delayed naturalistic recall task. For additional details see Figure S3. **c. Foreign language flashcards.** This panel is in the same format as Figure 2c, but it reflects performance on the delayed flashcards test. For additional details see Figure S4.



**Figure 4: Fitness measures.** **a. Recent measures.** Resting heart rate (HR) and daily step counts, averaged over the week prior to testing. **Baseline-normalized measures.** Resting heart rate and daily step counts averaged over the week prior to testing, divided by the average resting heart rate and step counts averaged over the preceding month. **Static measures.** Body mass index (BMI), body fat percentage, and weight (in kg). For more information see Figures S6, S7, S8, S9, and S10.

352 (see *Exploratory correlation analyses*). When all 10,000 randomly chosen subsets of participants ex-  
353 hibited correlations in the same direction (i.e., all positive correlations or all negative correlations),  
354 we report the  $p$ -value as  $p < 0.0001$ .

355 Several patterns emerged from these analyses. First, we found that participants' performance  
356 on the (within-task) immediate versus delayed memory tests from the free recall, naturalistic  
357 recall, and foreign language flashcards tasks were positively correlated ( $rs > 0.25, ps < 0.003$ ).  
358 This suggests that, within each of these tasks, similar processes or constraints may influence both  
359 short term and long term information retrieval. We also found reliable across-task correlations  
360 between participants' (immediate and delayed) performance on the free recall and foreign language  
361 flashcards tasks ( $rs > 0.3, ps < 0.03$ ).

362 A large number of fitness-related measures displayed reliable correlations (for a complete re-  
363 port, see Fig. S12). For example, body mass index (BMI) and weight were correlated ( $r = 0.91, p <$   
364  $0.0001$ ). Resting heart rate over the prior week was negatively correlated with recent low-to-  
365 moderate-intensity ("fat burn") cardiovascular activity levels ( $r = 0.70, p = 0.0004$ ). Participants'  
366 peak heart rates (averaged over the prior week) were also negatively correlated with recent in-  
367 creases in step counts and daily elevation gains ( $rs < -0.26, ps < 0.03$ ), where recent changes  
368 were defined as the average values over the seven days leading up to the test day divided by  
369 the average values over the preceding 30 days. Several demographic attributes (Fig. S13) dis-  
370 played trivial correlations (e.g., participants identifying as male never reported identifying as  
371 female, and so on). We also observed a negative correlation between reported stress and alertness  
372 ( $r = -0.44, p < 0.0001$ ), and positive correlations between the reported clarity of the instructions  
373 for all tasks ( $rs > 0.26, ps < 0.02$ ).

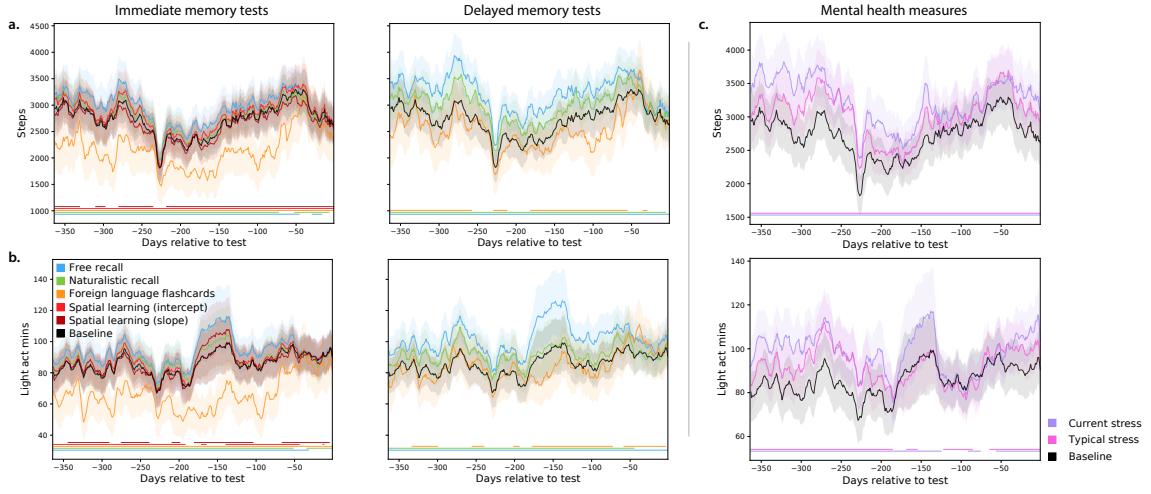
374 We also found reliable correlations between participants' fitness and demographic measures  
375 and their behaviors in different tasks (for a complete report, see Fig. S14). For example, recent  
376 low-to-moderate-intensity ("fat burn") cardiovascular activity was positively correlated with im-  
377 mediate ( $r = 0.38, p = 0.03$ ) and delayed ( $r = 0.38, p = 0.029$ ) recall performance on the naturalistic  
378 memory task. Recent increases in moderate-intensity ("cardio") activity over the prior 7 days  
379 (relative to the preceding 30 days) was also positively correlated with immediate naturalistic re-

Behavioral measures	Mental health measures									
	Anxiety or depression	High blood pressure	Bipolar	Hypothyroid	Unspecified medications	Recent head injury	Current stress	Typical stress	Current / typical stress	Alertness
Free recall (immediate)		-0.11	0.11	-0.16	0.16	-0.05				
Free recall (delayed)			0.20	-0.16	0.10	-0.13		0.17		
Naturalistic recall (immediate)	0.07	0.12	-0.25			0.05				
Naturalistic recall (delayed)	-0.06	0.12	-0.15	0.10	-0.10					
Foreign language flashcards (immediate)	0.11				0.09	0.17				
Foreign language flashcards (delayed)		0.15		0.15			-0.29		-0.15	
Spatial learning (intercept)	0.02		-0.14	-0.10						
Spatial learning (slope)	0.04		-0.03	0.04	0.13				-0.19	

**Figure 5: Memory performance differs according to mental health measures.** The reported values in the table reflect correlations between each behavioral measure and mental health measure. Only statistically reliable correlations ( $p < 0.05$ , corrected) are displayed. We used participants' mean recall accuracy to characterize performance on the free recall and foreign language flashcards tasks, and mean precision to characterize performance on the naturalistic recall tasks. We characterized performance on the spatial learning task using the (inverted and normalized) intercepts and slopes of linear regressions on mean estimation errors as a function of the number of studied shapes (also see Figs. 2, 3, S2, S3, S4, and S5). Typical and current stress levels were measured by self report. Mental health information was inferred using each participants' list of self-reported medications.

380 call performance ( $r = 0.48, p = 0.003$ ) and immediate recall performance on the foreign language  
381 flashcards task ( $r = 0.43, p = 0.048$ ). Recent high-intensity (“peak”) activity was positively corre-  
382 lated with performance on the spatial learning task ( $r = 0.34, p < 0.0001$ ), as were recent increases  
383 in high-intensity activity (prior 7 days versus preceding 30 days;  $r = 0.41, p = 0.01$ ). Mental  
384 health indicators, such as self-reported stress levels and medications were also associated with  
385 differences in memory (Figs. 5, S14). For example, self-reported stress levels at the time of test  
386 were negatively correlated with performance on the delayed memory test for the foreign language  
387 flashcards task ( $r = -0.29, p = 0.038$ ), whereas participants who were medicated for anxiety and  
388 depression tended to perform slightly (but reliably) *better* on the immediate memory test for the  
389 foreign language flashcards task ( $r = 0.11, p < 0.0001$ ).

390 The above analyses indicate that recent differences in fitness-related activity are associated with  
391 differences in memory performance and mental health measures. Although the analyses treated  
392 these measures on average or in aggregate, many of the measures we collected are dynamic. For  
393 example, the amount or intensity of physical exercise people engage in can vary over time, and  
394 so on. We wondered whether the dynamics of fitness-related measures might relate to memory  
395 performance and/or mental health measures. To this end, we carried out a series of reverse  
396 correlation analyses (see *Reverse correlation analyses*) to examine whether participants with different  
397 cognitive or mental health profiles also tended to display differences in fitness-related measures  
398 over time. In particular, we examined fitness data collected from participants’ Fitbit devices over the  
399 year prior to their test day in our study. Several example findings are summarized in Figure 6. We  
400 found that participants who performed well on the immediate and delayed free recall memory tests  
401 and on the naturalistic recall tests tended to be more active than participants who performed poorly  
402 on those tests (Figs. 6a, b; S15). Conversely, participants who performed well on the immediate  
403 and delayed foreign language flashcards tasks tended to be *less* active. These differences were  
404 present even a full year before the testing day. We also found substantial variability across people  
405 with different (self-reported) mental health profiles (Figs. 6c, S18). Due to small sample sizes of  
406 individuals exhibiting several mental health dimensions, it is difficult to distinguish generalizable  
407 trends from individual differences that one or two individuals happened to exhibit. However,



**Figure 6: Dynamics of physical activity varies with memory performance and mental health measures.** **a. Daily step counts.** Each timecourse is weighted by either performance on immediate recall tests (left panel) or on delayed recall tests (right panel). The black (baseline) timecourses display the (unweighted) average across all participants. **b. Daily duration (in minutes) of low-intensity physical activity.** Timecourses are displayed in the same format and color scheme as those in Panel A. Analogous timecourses for additional fitness-related measures may be found in Figures S15, S16, and S17. **c. Timecourses of physical activity, weighted by mental health measures.** The timecourses in each panel display the average daily step counts (top panel) or duration of low-intensity activity (bottom panel). The colored lines show average activity dynamics weighted by self-reported stress levels at the start of the experiment (purple) and self-reported “typical” stress levels (pink). The baseline curves (black) display the average across all participants (re-plotted in Panel C to illustrate scale differences across panels). Timecourses for additional mental health-related and fitness-related measures may be found in Figures S18, S19, and S20. Error ribbons in all panels denote the standard error of the mean. Horizontal lines below each panel’s timecourses denote intervals over which each weighted measure (color) differs from the unweighted baseline (via a paired sample two-sided  $t$ -test of the weighted mean values for each measure within a 30 day window around each timepoint; horizontal lines denote  $p < 0.05$ , corrected).

408 several large-sample-size trends emerged. For example, participants who reported higher levels  
409 of stress also tended to be slightly more physically active than participants who reported lower  
410 stress levels. We found analogous differences in other activity-related measures (Figs. S15 and S18),  
411 cardiovascular measures (Figs. S16 and S19), and sleep-related measures (Figs. S17 and S20). Taken  
412 together, the analyses suggest that cognitive and mental health differences are also associated with  
413 differences in the dynamics of physical health measures.

## 414 Discussion

415 After collecting a year's worth of fitness-tracking data from each of 113 participants, we ran each  
416 participant in a battery of memory tasks and had them fill out a series of demographic and mental  
417 health-related questions. We found that the associations between fitness-related activities, memory  
418 performance, and mental health were heterogeneous. Our results suggest that engagement in  
419 particular physical activities (e.g., differing in time relative to the test day, intensity, duration, etc.)  
420 is also reflected in participants' memory performance patterns across tasks and in participants'  
421 mental health attributes.

422 One important limitation of our study is that we cannot distinguish correlations between  
423 different measures from potential causal effects. For example, we cannot know (from our study)  
424 whether engaging in particular forms of exercise *causes* changes in memory performance or mental  
425 health, or whether (alternatively) people who tend to engage in similar forms of exercise also  
426 happen to exhibit similar memory and/or mental health profiles. In other words, an overlapping  
427 set of processes or person-specific attributes may lead someone to both form particular habits  
428 around exercise and display high or low performance on a given memory test. We do not know  
429 whether memory performance or aspects of mental health might be manipulated or influenced  
430 by changing the patterns of physical activity someone engages in. For this reason, we have been  
431 careful to frame our findings as correlations and associations, rather than to imply knowledge  
432 about a causal direction of our findings.

433 Although the present study cannot reveal causal effects, a large prior literature provides some

434 insight into potential causal effects by examining the neural and cognitive effects of a variety of  
435 exercise interventions [5, 14, 17, 34, 35, 37, 38]. A limitation of that prior work is that most of  
436 these studies examine how relatively short-term changes in exercise (e.g., on timescales of hours to  
437 days or, rarely, weeks to months) affect a cognitive performance on single task or aspect of mental  
438 health. The present study examines longer-term exercise (over a full year), and relates long-term  
439 exercise history to performance on a variety of tasks and to a variety of mental health dimensions.

440 To the extent that exercise *does* provide a non-invasive means of manipulating cognitive per-  
441 formance and mental health, our work may have exciting implications for cognitive enhancement.  
442 For example, one might imagine building a recommendation system that suggests a particular ex-  
443 ercise regimen tailored to improve a specific aspect of an individual’s cognitive performance (e.g.,  
444 the efficacy of a student’s study session for an upcoming exam) or mental health (e.g., reducing  
445 symptoms of anxiety before an important meeting). Just as strength training may be customized  
446 to target a specific muscle group, or to improve performance on a specific physical task, similar  
447 principles might also be applied to target specific improvements in cognitive fitness and mental  
448 health.

## 449 Acknowledgements

450 We acknowledge useful discussions with David Bucci, Emily Glasser, Andrew Heusser, Abigail  
451 Bartolome, Lorie Loeb, Lucy Owen, and Kirsten Ziman. Our work was supported in part by the  
452 Dartmouth Young Minds and Brains initiative, and by NSF grant number 2145172 to J.R.M. The  
453 content is solely the responsibility of the authors and does not necessarily represent the official  
454 views of our supporting organizations. This paper is dedicated to the memory of David Bucci, who  
455 helped to inspire the theoretical foundations of this work. Dave served as a mentor and colleague  
456 on the project prior to his passing.

<sup>457</sup> **Data and code availability**

<sup>458</sup> All analysis code and data used in the present manuscript may be found [here](#).

<sup>459</sup> **Author contributions**

<sup>460</sup> Concept: J.R.M. and G.M.N. Experiment implementation and data collection: G.M.N. Analyses:

<sup>461</sup> J.R.M., G.M.N., E.C., and P.C.F. Writing: J.R.M. with input from all authors.

<sup>462</sup> **Competing interests**

<sup>463</sup> The authors declare no competing interests.

<sup>464</sup> **References**

<sup>465</sup> [1] Bassey, E. J. and Ramsdale, S. J. (1994). Increase in femoral bone density in young women  
<sup>466</sup> following high-impact exercise. *Osteoporosis International*, 4:72–75.

<sup>467</sup> [2] Basso, J. C. and Suzuki, W. A. (2017). The effects of acute exercise on mood, cognition, neuro-  
<sup>468</sup> physiology, and neurochemical pathways: a review. *Brain Plasticity*, 2(2):127–152.

<sup>469</sup> [3] Brisswalter, J., Collardeau, M., and René, A. (2002). Effects of acute physical exercise charac-  
<sup>470</sup> teristics on cognitive performance. *Sports Medicine*, 32:555–566.

<sup>471</sup> [4] Callaghan, P. (2004). Exercise: a neglected intervention in mental health care? *Psychiatric and*  
<sup>472</sup> *Mental Health Nursing*, 11(4):476–483.

<sup>473</sup> [5] Chang, Y.-K., Chu, C.-H., Wang, C.-C., Wang, Y.-C., Song, T.-F., Tsai, C.-L., and Etnier, J. L.  
<sup>474</sup> (2015). Dose-response relation between exercise duration and cognition. *Medicine and Science in*  
<sup>475</sup> *Sports and Exercise*, 47:159–165.

- 476 [6] Chang, Y. K., Labban, J. D., Gapin, J. I., and Etnier, J. L. (2012). The effects of acute exercise on  
477 cognitive performance: a meta-analysis. *Brain Research*, 1453:87–101.
- 478 [7] Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared  
479 memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*,  
480 20(1):115.
- 481 [8] Chilibeck, P. D., Sale, D. G., and Webber, C. E. (2012). Exercise and bone mineral density. *Sports  
482 Medicine*, 19:103–122.
- 483 [9] Crane, J. D., MacNeil, L. G., and Tarnopolsky, M. A. (2013). Long-term aerobic exercise is  
484 associated with greater muscle strength throughout the life span. *The Journals of Gerontology:  
485 Series A*, 68(6):631–638.
- 486 [10] Deslandes, A., Moraes, H., Ferreira, C., Veiga, H., Silveira, H., Mouta, R., Pompeu, F. A. M. S.,  
487 Coutinho, E. S. F., and Laks, J. (2009). Exercise and mental health: many reasons to move.  
488 *Neuropsychobiology*, 59:191–198.
- 489 [11] Etnier, J. L., Nowell, P. M., Landers, D. M., and Sibley, B. A. (2006). A meta-regression to  
490 examine the relationship between aerobic fitness and cognitive performance. *Brain Research:  
491 Brain Research Reviews*, 52(1):119–130.
- 492 [12] Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern,  
493 D., Hamrick, J. B., and Chan, P. (2016). psiTurk: an open-source framework for conducting  
494 replicable behavioral experiments online. *Behavior Research Methods*, 48(3):829–842.
- 495 [13] Heusser, A. C., Fitzpatrick, P. C., and Manning, J. R. (2021). Geometric models reveal be-  
496 havioral and neural signatures of transforming naturalistic experiences into episodic memories.  
497 *Nature Human Behavior*, 5:905–919.
- 498 [14] Imboden, C., Gerber, M., Beck, J., Eckert, A., Pühse, U., Holsboer-Trachsler, E., and Hatzinger,  
499 M. (2019). Effects of aerobic exercise as add-on treatment for inpatients with moderate to severe  
500 depression on depression severity, sleep, cognition, psychological well-being, and biomarkers:

- 501 study protocol, description of study population, and manipulation check. *Frontiers in Psychiatry*,  
502 10(262):doi.org/10.3389/fpsyg.2019.00262.
- 503 [15] Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition*,  
504 24:103–109.
- 505 [16] Kahana, M. J. (2012). *Foundations of human memory*. Oxford University Press, New York, NY.
- 506 [17] Kamijo, K., Nishihira, Y., Higashiura, T., and Kuroiwa, K. (2007). The interactive effect of  
507 exercise intensity and task difficulty on human cognitive processing. *International Journal of*  
508 *Psychophysiology*, 65(2):114–121.
- 509 [18] Knuttgen, H. G. (2007). Strength training and aerobic exercise: comparison and contrast.  
510 *Journal of Strength and Conditioning Research*, 21(3):973–978.
- 511 [19] Layne, J. E. and Nelson, M. E. (1999). The effects of progressive resistance training on bone  
512 density: a review. *Medicine and Science in Sports and Exercise*, 31(1):25–30.
- 513 [20] Lazovic-Popovic, B., Zlatkovic-Svenda, M., Durmic, T., Djelic, M., Saranovic, D., and Zugic, V.  
514 (2016). Superior lung capacity in swimmers: some questions, more answers! *Revista Portuguesa*  
515 *de Pneumologia*, 22(3):151–156.
- 516 [21] Lindh, M. (1979). Increase of muscle strength from isometric quadriceps exercises at different  
517 knee angles. *Scandinavian Journal of Rehabilitation Medicine*, 11(1):33–36.
- 518 [22] Maiorana, A., O'Driscoll, G., Cheetham, C., Collis, J., Goodman, C., Rankin, S., Taylor, R.,  
519 and Green, D. (2000). Combined aerobic and resistance exercise training improves functional  
520 capacity and strength in CHF. *Journal of Applied Physiology*, 88(1565–1570).
- 521 [23] Mikkelsen, K., Stojanovska, L., Polenakovic, M., Bosevski, M., and Apostolopoulos, V. (2017).  
522 Exercise and mental health. *Maturitas*, 106:48–56.
- 523 [24] Morton, J. P., Kayani, A. C., McArdle, A., and Drust, B. (2009). The exercise-induced stress  
524 response of skeletal muscle, with specific emphasis on humans. *Sports Medicine*, 39:643–662.

- 525 [25] Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology: General*, 64:482–488.
- 526
- 527 [26] Palacios-Filardo, J., Udkakis, M., Brown, G. A., Tehan, B. G., Congreve, M. S., Nathan, P. J.,  
528 Brown, A. J. H., and Mellor, J. R. (2021). Acetylcholine prioritises direct synaptic inputs from  
529 entorhinal cortex to CA1 by differential modulation of feedforward inhibitory circuits. *Nature Communications*, 12(5475):doi.org/10.1038/s41467-021-25280-5.
- 530
- 531 [27] Paluska, S. A. and Schwenk, T. L. (2000). Physical activity and mental health. *Sports Medicine*,  
532 29(3):167–180.
- 533
- 534 [28] Pollock, M. L., Franklin, B. A., Balady, G. J., Chaltman, B. L., Fleg, J. L., Fletcher, B., Limacher,  
535 M., na, I. L. P., Stein, R. A., Williams, M., and Bazzarre, T. (2000). Resistance exercise in individuals  
with and without cardiovascular disease. *Circulation*, 101:828–833.
- 536
- 537 [29] Raglin, J. S. (1990). Exercise and mental health. *Sports Medicine*, 9:323–329.
- 538
- 539 [30] Rogers, M. A. and Evans, W. J. (1993). Changes in skeletal muscle with aging: effects of  
540 exercise training. *Exercise and Sport Sciences Reviews*, 21:65–102.
- 541
- 542 [31] Roman, M. A., Rossiter, H. B., and Casaburi, R. (2016). Exercise, ageing and the lung. *European Respiratory Journal*, 48:1471–1486.
- 543
- 544 [32] Schiaffino, S., Dyar, K. A., Ciciliot, S., Blaauw, B., and Sandri, M. (2013). Mechanisms regulating  
545 skeletal muscle growth and atrophy. *The febs Journal*, 280(17):4294–4314.
- 546
- 547 [33] Shoemaker, J. K., Halliwill, J. R., Hughson, R. L., and Joyner, M. J. (1997). Contributions of  
548 acetylcholine and nitric oxide to forearm blood flow at exercise onset and recovery. *Vascular Physiology*, 273(5):2388–2395.
- 549
- 550 [34] Sinha, N., Berg, C. N., Yassa, M. A., and Gluck, M. A. (2021). Increased dynamic flexibility in  
the medial temporal lobe network following an exercise intervention mediates generalization  
of prior learning. *Neurobiology of Learning and Memory*, 177:107340.

- 549 [35] Suwabe, K., Hyodo, K., Byun, K., Ochi, G., Yassa, M. A., and Soya, H. (2017). Acute moderate  
550 exercise improves mnemonic discrimination in young adults. *Hippocampus*, 27(3):229–234.
- 551 [36] Taylor, C. B., Sallis, J. F., and Needle, R. (1985). The relation of physical activity and exercise  
552 to mental health. *Public Health Reports*, 100(2):195–202.
- 553 [37] Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychologica*,  
554 112(3):297–324.
- 555 [38] Vidoni, E. D., Johnson, D. K., Morris, J. K., Van Sciver, A., Greer, C. S., Billinger, S. A.,  
556 Donnelly, J. E., and Burns, J. M. (2015). Dose-response of aerobic exercise on cognition: A  
557 community-based, pilot randomized controlled trial. *PLoS One*, 10(7):1–13.
- 558 [39] Wilmore, J. H. and Knuttgen, H. G. (2003). Aerobic exercise and endurance. *The Physician and  
559 Sportsmedicine*, 31(5):45–51.
- 560 [40] Ziman, K., Heusser, A. C., Fitzpatrick, P. C., Field, C. E., and Manning, J. R. (2018). Is  
561 automatic speech-to-text transcription ready for use in psychological experiments? *Behavior  
562 Research Methods*, 50:2597–2605.